

Review

# Metric-based simulation training to proficiency in medical education:- What it is and how to do it

Anthony G Gallagher

Accepted 25 July 2012

## ABSTRACT

High profile error cases and reduced work hours have forced medicine to consider new approaches to training. Simulation-based learning for the acquisition and maintenance of skills has a growing role to play. Considerable advances have been made during the last 20 years on how simulation should be used optimally. Simulation is also more than a technology learning experience for supplanting the traditional approach of repeated practice. Research has shown that simulation works best when it is integrated into a curriculum. Learning is optimal when trainees receive metric-based feedback on their performance. Metrics should unambiguously characterize important aspects of procedure or skill performance. They are developed from a task analysis of the procedure or skills to be learned. The outcome of the task analysis should also shape how the simulation looks and behaves. Metric-based performance characterization can be used to establish a benchmark (i.e., a level of proficiency) which trainees must demonstrate before training progression. This approach ensures a more homogeneous skill-set in graduating trainees and can be applied to any level of training. Prospective, randomized and blinded clinical studies have shown that trainees who acquired their skills to a level of proficiency on a simulator in the skills laboratory perform significantly better *in vivo* in comparison to their traditionally trained colleagues. The Food and Drug Administration in the USA and the Department of Health in the UK have candidly indicated that they see an emergent and fundamental role for simulation-based training. Although a simulation-based approach to medical education and training may be conceptually and intellectually appealing it represents a paradigm shift in how doctors are educated and trained.

## BACKGROUND

During the Annual meeting of the American Surgical Association (ASA) in April 2002 at the Homestead (VA, USA) researchers from Queen's University Belfast and Yale University (USA) reported the results from the first prospective, randomized, double-blinded trial of virtual reality (VR) training for the operating room (OR). Surgical residents randomized to the VR training arm of this study subsequently performed the actual dissection of the gallbladder from the liver bed portion of a laparoscopic cholecystectomy (LC) 26% faster than traditionally trained surgical residents and made six times fewer objectively assessed intra-operative errors<sup>1</sup>. This study (or VR to OR as it has become known<sup>2</sup>) was important because for the

first time it unambiguously demonstrated the potential of (VR) simulation as a powerful training methodology for the acquisition of procedural skills outside the OR which directly impacted on *in vivo* OR performance. Previous studies had compared VR training to no, or traditional approaches to training (i.e. on real patients), but these studies were conducted wholly in the skills laboratory and not on real patients.<sup>3-7</sup>

Although simulation had been used in other industries such as aviation for decades<sup>8</sup> and in medical disciplines such as anaesthetics for many years<sup>9</sup> its potential as a training device in surgery and procedure-based medicine was not taken seriously until a series of high profile and impactful events forced medicine and surgery to consider a new way of training. High profile error cases in the UK<sup>10</sup> and the USA<sup>11</sup> as well reduced work (and training) hours<sup>12, 13</sup> forced medicine to consider new training paradigms. Bad experiences following the introduction of minimally invasive surgery (MIS) ensured that surgeons were already sensitised to the need to improve training. The introduction of MIS, particularly LC, was accompanied by an increased frequency of complications, many life-threatening, particularly during the early experiences.<sup>14</sup> That these problems could occur when experienced surgeons, well versed in open techniques and with knowledge of anatomy and pitfalls embraced new techniques, heightened concerns about the training of novices who lacked such a background in open surgery. But the agenda was now set, surgery needed to develop new methods for training the novice in surgical techniques in general and for training experienced surgeons in the newer techniques.<sup>15</sup>

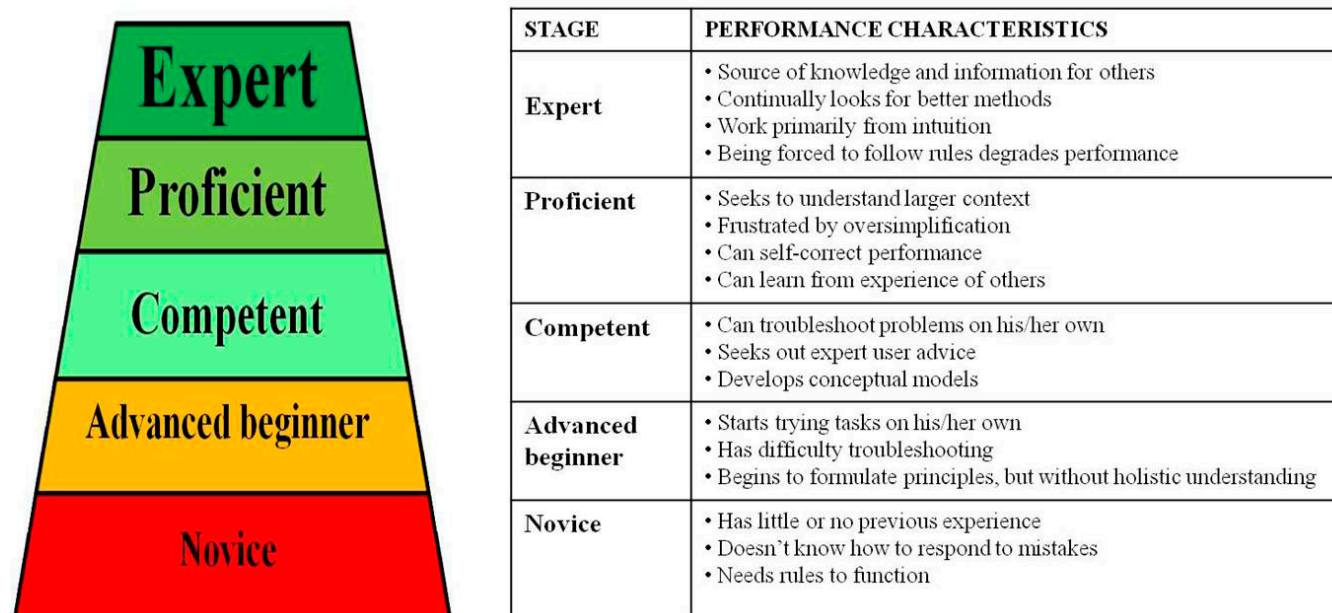
Dr. Richard Satava, a US army general surgeon first proposed the idea of VR simulation for the acquisition of surgical skills in the early 1990s whilst on secondment to the Defence Advanced Research Projects Agency (DARPA).<sup>16</sup> Until the VR to OR study results were reported his proposal was at best considered 'eccentric' and VR simulation a technology of marginal significance. In March 2002 at a closed door meeting in Boston College, hosted by Dr. Gerry Healy (President of the American College of Surgeons 2007 – 2008) and the American College of Surgeons (ACS) and in light of the imminent VR to OR study report at the ASA, the decision was

---

School of Medicine, Brookfield Health Sciences Complex, College Road,  
University College Cork, Cork, Ireland

anthonyg.gallagher@btinternet.com

Correspondence to Professor Anthony G Gallagher

Fig 1. Dreyfus & Dreyfus Figure<sup>23</sup>

taken for the ACS to 'champion surgical simulation'.<sup>17</sup> The impact of this brave and enlightened decision was enormous. True to its word the ACS championed simulation based training which has led to the establishment of Accredited Education Institutes (within the USA and globally) of which simulation is an important pedagogical component.<sup>18</sup>

### WHAT IS SIMULATION?

Simulations have been given many definitions over the years but it is fair to say that 'simulation' is usually thought of as VR. In 1993 Satava<sup>16</sup> originally proposed that "Virtual reality [simulation] is a fully three-dimensional computer-generated "world" in which a person can move about and interact as if he actually were in this imaginary place."... "A world can be anything from a kitchen to an automobile to an abdomen -- anything which can be drawn can be experienced three dimensionally". In contrast, high fidelity simulations in disciplines such as anaesthetics utilized very realistic materials and equipment to represent the task(s) that the candidate must perform<sup>19</sup>. Anaesthetic simulations centred on computerized, interactive, life-sized manikin that could be programmed to provide realistic patient responses and outcomes. Low fidelity simulations would be ones in which the candidate is presented with a verbal description of a hypothetical work situation and then asked to describe how he/she would deal with the situation rather than having the candidate perform the actions he/she would take.<sup>20</sup> Thus, simulation fidelity has been construed by trainers and educators as the degree of similarity (and apparent technical sophistication) of the simulation to the real world situation that was being trained. Inanimate simulation tasks such as animal tissue would be low fidelity and increasing sophistication and/or face validity of the simulation represented increasing simulation fidelity. This issue will be returned to when the concept of metrics is discussed. It suffices to state at this point that both of these views of simulation fidelity are now incomplete in light of a more sophisticated and comprehensive theoretical understanding of what a simulation is and how it is configured

and implemented for the efficient and effective learning of skilled performance.

### WHAT IS SKILL?

One of the most important functions of a simulation is to facilitate the effective and efficient training of skill outside the clinical situation thus minimising the risk to the patient from at least part of the novice's learning curve. But what is skill? Failure by medicine to explicitly answer this question has been one of the major impediments to the development of good simulations and simulation-based training. When United States Supreme Court Justice Potter Stewart was asked describe his threshold test for obscenity in the case of *Jacobellis v. Ohio* (1964) he infamously ruled that 'he knew it when he saw it'.<sup>21</sup> Similarly, most doctors have an opinion as to what skill is but few can robustly define it. A parsimonious definition of skill might be 'it is what skilled individuals do'. However, this definition does not advance a specific testable model of skilled performance that could be used to characterise performance. In contrast, psychologists have tackled the same problem by subjecting the 'skill' to be characterised to a detailed task analysis and then operationally defining (not describing) important aspects of performance which constitute skill.<sup>15,22</sup> They then quantitatively validate whether their characterisation fits with what is known about the skill they have analysed. Do more skilled individuals perform better on their assessments than less skilled or experienced individuals (construct validity)? Do individuals who perform well on their evaluations also perform well on a variety of similar and related tasks (concurrent validity)? Do their assessments predict future skilled performance (predictive validity)? These task-analysis derived characterizations of skilled performance do not have to capture every aspect of performance but should at least allow for ordinal differentiation between different levels of performance as described by Dreyfus and Dreyfus<sup>23</sup> and summarised in **Figure 1**<sup>15</sup>.

TABLE 1:  
*Metric Errors*

Metric errors/criteria of injury assessment	Operational Definition
Procedure <b>START</b>	first contact of diathermy with tissue
Procedure <b>END</b>	last attachment is divided
FAILURE TO PROGRESS	No progress made in excising the gallbladder for an entire minute of the dissection. Dealing with the consequences of a predefined error represents lack of progress if no progress is made in excising the gallbladder during this period.
GALLBLADDER INJURY	There is gallbladder wall perforation with or without leakage of bile. Injury may be incurred with either hand.
LIVER INJURY	Necessitates capsule penetration and may have bleeding.
BURN NONTARGET TISSUE	Any application of electrocautery to non-target tissue with the exception of the final part of the fundic dissection where some current transmission may occur.
TEARING TISSUE	Uncontrolled tearing of tissue with the dissecting or retracting instrument.
INCORRECT PLANE OF DISSECTION	The dissection is conducted outside the recognized plane between the gallbladder and the liver (i.e. in the sub-mucosal plane on the gallbladder, or sub-capsular plane on the liver).
INSTRUMENT OUT OF VIEW	The dissecting instrument is placed outside the field of view of the telescope such that its tip is un-viewable and can potentially be in contact with tissue. No error will be attributed to an incident of a dissecting instrument out of view as the result of a sudden telescope movement.
ATTENDING TAKEOVER	The supervising attending surgeon takes the dissecting instrument (right hand), or retracting instrument (left hand) from the resident and performs a component of the procedure.

## WHAT ARE METRICS?

Based on the task analysis process outlined above, the units of performance that have been identified (and validated) as integral to skilled task performance are the metric units of task execution. This means that these performance units should be used to define and shape the configuration of any simulation developed to train skilled task performance. Metric units must be unambiguously defined so that they can be scored as occurring or not occurring. These metric units should capture the essence of procedure performance and might include the steps that the procedure should be performed in, the instruments used and what should be done with them. Crucially, the metrics should also describe for each procedure step what not be done thus characterising performance that deviates from optimal performance (or errors).<sup>24</sup> Metric errors are some of the most important performance units for simulation based training.<sup>15</sup> Training should concentrate on what should be done and the order in which it should be done but it should also target performance errors for at least reduction, preferably elimination. This means that operational definitions of performance units or metrics need to be unambiguous. For example, **Table 1** shows the operational definitions of metrics for the dissection of the gallbladder from the liver bed portion of a LC including a

start and end point procedure markers.<sup>1</sup> They unambiguously 'define' rather than describe when each metric error has occurred. This approach considerably facilitates the reliable scoring of metric-based performance units across a variety of functions from skills training<sup>25-28</sup> at different experience levels.<sup>29, 30</sup> It has also been shown that this approach works well as part of the process for selection into higher training<sup>31</sup> and considerably enhances assessment reliability levels in comparison to Likert-scale assessments.<sup>32</sup>

In discussions with different groups of physicians and surgeons from around the world there appears to be a consensus appears that reaching agreement on performance metrics is all but impossible. This may however, only apply to agreement on 'everything'. The majority of doctors very experienced and proficient in the performance a specific procedure can very quickly identify and agree what should be done, how it should be done and what most certainly should not be done for most parts of a procedure. The problem is that doctors rarely think about procedures in that level of detail. Practitioners who are 'proficient' in the performance of a procedure (with average to good outcomes) will already exhibit many if not all of the important performance characteristics to perform the procedure well. They have



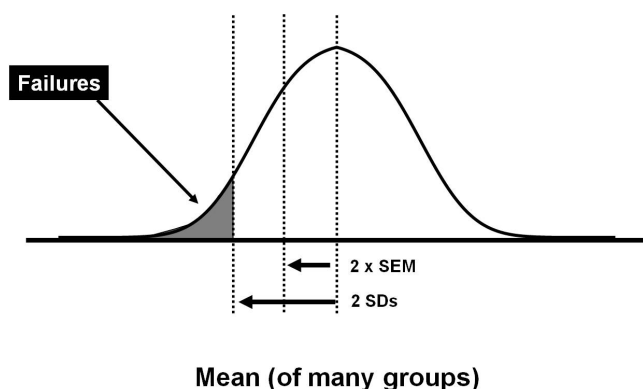


Fig 2a. the Wijen method of competency assessment

however, automated to many of these and how it is they are performed (much like the complex skills required to ride a bicycle). A primary function of the task analysis process is to identify and define what these performance characteristics are. This should be done initially for a 'reference procedure', i.e., a straightforward procedure that can be performed without complications or deviations under ideal circumstances. (An optimal approach to learning should ensure that trainees are capable of performing an uncomplicated procedure before they have to deal with procedure variations). The task analysis should seek consensus (not necessarily agreement) between procedure experts on the characteristics of the reference procedure and instruct them to characterise the reference procedure and not unusual or interesting variants of it. Procedure performance should be guided by a) professional guidelines, b) manufacturer guidelines on device usage and c) results from empirical studies. In the absence of a consensus between the experts on the items (a) to (c) above then individual procedure practices that they may have developed from years of practice wisdom should be employed. Errors are defined as procedure actions which deviate from optimal practice and are not necessarily bad but are potentially unsafe. Critical errors in contrast are procedure actions which are most certainly unsafe but may not always lead to a bad outcome.<sup>15</sup> The underlying philosophy of this approach is that bad outcomes do not happen by accident but usually from the coalescence of deviations from optimal procedure protocol.<sup>33</sup> The task analysis stage of the development of a simulation is crucial as metrics are the fundamental building blocks of a good training program. Metrics thus define how the simulation should be characterised and performed by the trainee and must afford the opportunity for meaningful performance assessment. Assuming that the metric identification and definition process, simulation operationalization and implementation goes well these performance characteristics should be easily validated, as distinguishing between experts and novices (i.e., construct validity) and predictive of acquired skills post training (predictive validity). Other validation processes are necessary but these two are probably the most important for training purposes. The construct validity study will inform the training process which metrics best distinguish between experienced/expert and novice performances and will guide the skills benchmarking process or 'proficiency level' which trainees should acquire before progressing to *in vivo* practice.<sup>15, 34</sup>

## SIMULATION DEFINED

Equipped with an understanding of the importance of metrics for the characterization and configuration of a simulation, simulation can thus be defined as i) an artificially created or configured 'learning' situation that allows for the practice or rehearsal of all or salient aspects of a procedure. The artificial learning situation must ii) provide the span of appropriate sensory responses to learner physical actions that are behaviourally consistent with what would be experienced in real life (including the opportunity to enact both appropriate and inappropriate learner actions (i.e., errors)). The simulation should also afford the opportunity to iii) perform the procedure iv) in the same order and v) with the same devices that the procedure would normally be performed. Crucially, it should also afford vi) reliable and valid metric-based assessment of performance. Assessments must at a minimum vii) allow summative but preferably formative feedback on procedure performance proximate to task execution, particularly for metric errors.

## WHAT IS PROFICIENCY-BASED TRAINING?

Traditionally medicine has trained doctors to be 'competent' and this fact has been reinforced by high profile medical errors cases, e.g., The Bristol Case.<sup>10</sup> However, what is probably less well known outside of (academic) medicine is that this level of competency is in fact 'minimum' competence. Indeed over the years medicine has developed robust statistical processes for the definition of minimum competency levels (Figure 2a).<sup>35</sup> As shown in Figure 3a competency definitions can vary considerably between and even within institutions using these statistical methods. The concern that many surgeons, interventional cardiologists and radiologists and other procedurists express is that this level of skill may be too low. Also of concern to them is the variability in competence thresholds. They also express concern about the failure of medicine to define a measured skill level that is objective, transparent and fair. Furthermore, they believe that training practitioners to a level of proficiency<sup>23</sup> may be a more conservative but superior approach. They are however faced with the same problem of benchmark definition.

The "I know it when I see it" approach does not work well for understanding and operationally defining what skill is but, it will serve more than adequately for helping to define a proficiency benchmark, i.e., proficiency is the level of performance exhibited by proficient individuals. Thus, validated metric-based simulations can serve as benchmarking devices. The philosophy being that individuals who are very experienced at performing a specific task or procedure e.g., a consultant/master surgeon, are at least competent, probably proficient and some will be expert.<sup>15, 34</sup> This means that a level of proficiency benchmarked on the mean performance level of these individuals on the validated metric-based simulation is a fairly conservative criterion level. Furthermore, it has considerable face validity. Rather than benchmarking on some abstract performance level reached by consensus in a committee the training pass level is defined on the performance levels of individuals who are actually very experienced at performing the procedure clinically. This does not imply that individuals who demonstrate this benchmark in training have the same 'performance capacity' as the individuals on whom the level was established. It simply means that they were

able to demonstrate that performance level (e.g., technical or knowledge or both) on consecutive trials. They will not (yet) have acquired the procedural wisdom of the experienced operator. This approach to training ensures a much more heterogeneous skill set from trainees as ALL trainees must reach AT LEAST the proficiency level (on two consecutive trials) before progressing to *in vivo* practice (**Figure 2b**). The other advantage of this approach (summarized in **Figure 3**) is that it eliminates the issue of number of training trials or time in training before progressing. It also puts the onus on the trainee to demonstrate proficiency before progressing. The onus on the trainer is to provide the facilities and access to training to demonstrate proficiency. The results from this approach to training skills have shown that proficiency-based progression trainees significantly outperform traditionally trained doctors,<sup>1, 25, 36</sup> even for advanced procedures.<sup>28</sup>

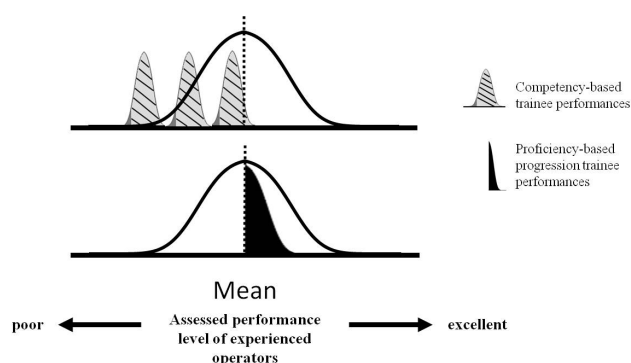


Fig 2b. Competency Vs Proficiency

### WHAT IS THE DIFFERENCE BETWEEN REPEATED AND DELIBERATE PRACTICE?

Medicine is currently grappling with the mechanics of a proficiency-based progression approach to training (e.g., development of metrics, metric validation, proficiency definition etc).<sup>15</sup> It circumvents some very sticky problems such as how to define competency? If an individual has demonstrated a proficiency benchmark they must by default have demonstrated competency as proficiency is a more advanced skill level as proposed by Dreyfus and Dreyfus.<sup>23</sup> It has also forced medicine to define precisely what is meant by 'skill' (for a reference approach at least). Another benefit from the process of defining skill and deviations from optimal performance, i.e., metrics, has been the evolution of an understanding how metrics could be used to define, shape and configure a specific simulation. Thus a simulation becomes a vehicle for the delivery of metric-based training rather than some abstract entity. Furthermore, metric-based simulation training can be configured in such a way as to make training much more efficient and effective. Traditionally skills in medicine have been acquired through repeated practice in the clinical setting. How quickly they learned depended on how frequently they encountered the learning situation, whether they have the opportunity to implement or practice their skills and whether their supervisor had the opportunity (or inclination) to use the setting as a training occasion. Thus, the learning accrued from such opportunities could be very variable. Simulation affords the trainee to acquire their skills in the simulation laboratory in a much more systematic way. It also means that each time the trainee engages with a metric-

based simulation their learning can be optimized. Each time the trainee performs the procedure they will receive feedback on their performance. Summative feedback at the end of the procedure will facilitate learning but metric-based formative feedback on performance is a much more powerful aid. This will inform trainees if they are performing the procedure incorrectly or in the wrong order and it should let them know if they are using the devices inappropriately. Trainees should receive this feedback proximate to the performance error. This approach to learning is called deliberate practice.<sup>37</sup>

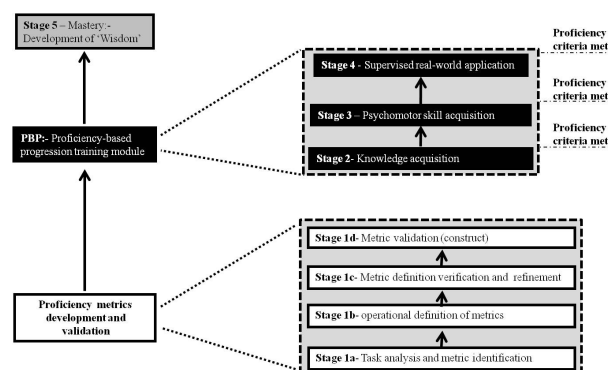


Fig 3. the Proficiency-based progression method

The extended, structured and motivated practice by trainees first described Ericsson, Krampe and Tesch-Römer<sup>37</sup> were important explanations of learning skilled performance but one crucial aspect of their accounts to training was missing, i.e., performance feedback. In addition to engaging in deliberate practice the trainee musicians studied by Ericsson and colleagues also had access to immediate feedback from their tutor. They would have been informed by their tutor (or recognised themselves) when they played a note wrong. Engaging in frequent practice sessions would mean that the trainee had ample opportunity to practice and rehearse their playing with equivalent opportunities to correct and improve performance. Implicit to this process and presumed performance improvement (i.e., learning) is the concept of performance feedback. The same is true for the acquisition of skill in medicine. This approach to learning is more efficient and effective than the traditional approach of repeated practice and is made possible with metric-based simulations. The pre-requisites for this approach to simulation training are the identification and definition of optimal and sub-optimal performance by procedural experts and the capability of the simulation to model the procedure and operator performance interaction in real-time. It is then simply a matter of benchmarking proficiency using the actual performance of proficient doctors. Thus the experienced doctor is the starting point for the development of a simulation and the quality assurance of it by benchmarking on their performance thus circumventing some potentially complicated, convoluted and thorny issues such as what is competence or how is it defined?

### THE FUTURE (NOW)

Metric-based simulation ensures that training sessions are more than just simulated clinical 'experiences'. It ensures that there is no ambiguity about the progress of training for the trainee (and the trainer). Simulation-based training in medicine now has a quantitatively validated

clinical function. This foothold is considerably better than even a decade ago. So much so that the Food and Drug Administration in the USA now requires simulation based training as part of device approval<sup>38, 39</sup> and the Department of Health in the UK have issued guidance which appears to suggest that trainees should not be performing a procedure on a real patient the first time they perform the procedure.<sup>40</sup> Implementation of simulation based training will continue to evolve but for certain, simulation based training will form a fundamental part of acquisition and maintenance of skills in medicine and healthcare. In time it will also (probably) be part of the General Medical Councils reaccreditation process. It is therefore essential that medicine has a thorough understanding of what simulation is and the imperative of metrics. Quantitative characterization that has been validated for a proficiency-based progression training function is not a million miles away from quantitative assessment of performance by experienced doctors, i.e., re-validation.

## SUMMARY

Proficiency-based progression training works and it works because of well proven principles and practices of learning.<sup>15</sup> To ensure the optimal effectiveness of a proficiency-based progression training program does not require a radical change in the current curriculum content. However, what it does require is a radical change in how that curriculum is delivered and implemented. Simulation is very powerful training tool for the delivery of deliberate practice coupled to formative and summative metric-based feedback on performance. In the absence of computer generated simulation, formative metrics on training performance need to be delivered by a trainer who is very experienced at performance assessment. A training program that has a clear end point (i.e., level of proficiency) must provide the facilities and opportunities for the learner to meet the level of proficiency. A deliberate practice training regime affords the opportunity for independent pacing of skill acquisition; a coherent curriculum with appropriately sequenced learning material and a variety of learning experiences. Although this approach to medical education and training may be conceptually and intellectually appealing it represents a paradigm shift in how doctors are educated and trained.

## REFERENCES

1. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, *et al.* Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg.* 2002;**236**(4):458-63.
2. Seymour NE. VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World J Surg.* 2008;**32**(2):182-8.
3. Jordan JA, Gallagher AG, McGuigan J, McGlade K, McClure N. A comparison between randomly alternating imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. *Am J Surg.* 2000;**180**(3):208-11.
4. Jordan JA, Gallagher AG, McGuigan J, McClure N. Virtual reality training leads to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons. *Surg Endosc.* 2001;**15**(10):1080-4.
5. Taffinder N, Sutton C, Fishwick R, McManus I, Darzi A. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Stud Health Technol Inform.* 1998;**50**:124-30.
6. Gallagher AG, McClure N, McGuigan J, Crothers I, Browning J. Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). *Endoscopy.* 1999;**31**(4):310-3.
7. Gallagher AG, Hughes C, Reinhardt-Rutland AH, McGuigan J, McClure N. A case-control comparison of traditional and virtual-reality training in laparoscopic psychomotor performance. *Minim Invasive Ther Allied Technol.* 2000;**9**(5):347-52.
8. Salas E, Bowers CA, Rhodenizer L. It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *Int J Aviat Psychol.* 1998;**8**(3):197-208.
9. Cooper JB, Taqueti VR. A brief history of the development of mannequin simulators for clinical education and training. *Qual Saf Health Care.* 2004;**13**(Suppl 1):i11 - i8. Erratum in *Qual Saf Health Care* 2005; **14**(1):72.
10. Smith R. All changed, changed utterly. *Br Med J.* 1998;**316**(7149):1917-8.
11. Kohn LT, Corrigan JM, Donaldson MS. To err is human. Building a safer health system. Volume 6. Washington DC: National Academy Press; 2000. p 196-7.
12. Beall DP. The ACGME institutional requirements: what residents need to know. *JAMA.* 1999;**281**(24):2352.
13. Lowry J, Cripps J. Results of the online EWTD trainee survey. *Bull R Coll Surg Engl.* 2005;**87**(3):86-7.
14. The Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. *New Engl J Med.* 1991;**324**(16):1073 - 8.
15. Gallagher AG, O'Sullivan GC. Fundamentals of surgical simulation; principles and practice. London: Springer Verlag; 2011.
16. Satava RM. Virtual reality surgical simulator. The first steps. *Surg Endosc.* 1993;**7**(3):203-5.
17. Healy GB. The college should be instrumental in adapting simulators to education. *Bull Am Coll Surg.* 2002;**87**(11):10-1.
18. Pellegrini CA, Sachdeva AK, Johnson KA. Accreditation of education institutes by the American College of Surgeons: a new program following an old tradition. *Bull Am Coll Surg.* 2006;**91**(3):8- 2.
19. Havighurst LC, Fields LE, Fields CL. High versus low fidelity simulations: does the type of format affect candidates performance or perceptions. Proceedings from the 27<sup>th</sup> annual IPMAAC conference on personnel assessment, 2003 June 22-25. Baltimore, USA: International Personnel Assessment Council [IPAC]; 2003. Available online from: <http://www.ipacweb.org/conf/03/havighurst.pdf>. Last accessed August 2012.
20. Motowidlo SJ, Dunnette MD, Carter GW. An alternative selection procedure: The low-fidelity simulation. *J Appl Psychol.* 1990;**75**(6):640-7.
21. Gewirtz P. On "I Know It When I See It". *Yale Law J.* 1996;**105**(4):1023-47.
22. Andersen DK. How can educators use simulation applications to teach and assess surgical judgment? *Acad Med.* 2012;**87**(7):934-41.
23. Dreyfus HL, Dreyfus SE, Athanasiou T. *Mind over machine: the power of human intuition and expertise in the era of the computer.* New York: Free Press; 1986.
24. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Andersen DK, Satava RM. Analysis of errors in laparoscopic surgical procedures. *Surg Endosc.* 2004;**18**(4):592-5.
25. McClusky DA, Gallagher AG, Ritter EM, Lederman AB, Van Sickel MB, Smith CD. Virtual reality training improves junior residents' operating room performance: Results of a prospective, randomized,



- double-blinded study of the complete laparoscopic cholecystectomy. *J Am Coll Surg.* 2004;**199**(Suppl 3):73.
26. Van Sickle K, Smith B, McClusky DA, Baghai M, Smith CD, Gallagher AG. Evaluation of a tensiometer to provide objective feedback in knot-tying performance. *Am Surg.* 2005;**71**(12):1018-23.
27. Van Sickle KR, Gallagher AG, Smith CD. The effect of escalating feedback on the acquisition of psychomotor skills for laparoscopy. *Surg Endosc.* 2007;**21**(2):220-4.
28. Van Sickle K, Ritter EM, Baghai M, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. *J Am Coll Surg.* 2008;**207**(4):560-8.
29. Neary PC, Boyle E, Delaney CP, Senagore AJ, Keane FB, Gallagher AG. Construct validation of a novel hybrid virtual-reality simulator for training and assessing laparoscopic colectomy; results from the first course for experienced senior laparoscopic surgeons. *Surg Endosc.* 2008 Oct;**22**(10):2301-9.
30. Nicholson WJ, Cates CU, Patel AD, Niazi K, Palmer S, Helmy T, et al. Face and content validation of virtual reality simulation for carotid angiography: results from the first 100 physicians attending the Emory NeuroAnatomy Carotid Training (ENACT) program. *Simul Healthc.* 2006;**1**(3):147-50.
31. Gallagher AG, Neary P, Gillen P, Lane B, Whelan A, Tanner WA, et al. Novel method for assessment and selection of trainees for higher surgical training in general surgery. *Aust N Z J Surg.* 2008;**78**(4):282-90.
32. Gallagher AG, O'Sullivan GC, Leonard G, Bunting BP, KJ. M. OSATS and checklist scales reliability compared for high stakes assessments. *Aust N Z J Surg.* In Press.
33. Reason J. *Human error*. Cambridge, New York: Cambridge University Press; 1990.
34. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg.* 2005 Feb;**241**(2):364-72.
35. Cusimano MD. Standard setting in medical education. *Acad Med.* 1996;**71**(Suppl 10):S112.
36. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McCluskey, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Amer J Surg.* 2007;**193**(6):797-804.
37. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev.* 1993;**100**(3):363-406.
38. Gallagher AG, Cates CU. Approval of virtual reality training for carotid stenting: what this means for procedural-based medicine. *JAMA.* 2004;**292**(24):3024 -6.
39. U.S. Department of Health and Human Services. Draft guidance for industry and food and drug administration staff. Applying human factors and usability engineering to optimize medical device design. Silver Spring, MD 20993: U.S Department of Health and Human Services. Food and Drug Administration Center for Devices and Radiological Health; 2011. Available online from: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM259760.pdf>. Last accessed August 2012.
40. Department of Health. Framework for technology enhanced learning. Best practice guidance. London: Department of Health; 2011. Available online from: [http://www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/documents/digitalasset/dh\\_131061.pdf](http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_131061.pdf). Last accessed August 2012.